

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application.

Listing of Claims:

1. **(Currently Amended)** A method of providing pattern recognition, said method comprising the steps of:

inputting a speech pattern;

providing minimum Bayes error feature selection via transforming the input pattern to provide a set of ~~at least one feature~~ features for a classifier which classifies into classes, wherein there is only one feature space transformation for all classes; and

providing final features to the classifier, wherein the classifier provides a final output classification result;

said transforming step comprising the step of directly minimizing the probability of subsequent misclassification in a projected space of at least one feature ~~in the~~ classifier; said direct minimizing step comprising:

performing a full-covariance gaussian clustering of input records for every class;

developing an objective function by way of means, covariances and priors,
wherein said objective function either:

maximizes an average pairwise divergence and relates it to Bayes error; or

directly minimizes an upper bound on Bayes error; and

optimizing the objective function through gradient decent, wherein all dimensions of a matrix are optimized via optimizing the objective function;

wherein the optimizing is carried out ~~in an unconstrained manner~~ over all possible matrices; and

wherein the objective function is initialized with an LDA matrix;

upon convergence of the optimization, transforming all the records into $y = \theta x$ to produce the at least one final feature;

wherein said pattern recognition is speech recognition.

2. **(Cancelled)**

3. **(Cancelled)**

4. **(Original)** The method of Claim 1, further comprising the step of querying whether the optimized objective function converges.

5. **(Original)** The method according to Claim 4, further comprising the step of repeating said optimizing step if the optimized objective function does not converge.

6. **(Cancelled)**

7. **(Currently Amended)** An apparatus for providing pattern recognition, said apparatus comprising:

an input interface for inputting a speech pattern;

a transformer for providing minimum Bayes error feature selection via transforming the input pattern to provide a set of ~~at least one feature~~ features for a classifier which classifies into classes, wherein there is only one feature space transformation for all classes; and

a classifier for producing a final output classification result upon being provided final features;

said transformer being adapted to directly minimize the probability of subsequent misclassification in a projected space of ~~the~~ at least one feature in the classifier;

said transformer further being adapted to:

perform a full-covariance gaussian clustering of input records for every class;

develop an objective function by way of means, covariances and priors, wherein said objective function either:

maximizes an average pairwise divergence and relates it to Bayes error; or

directly minimizes an upper bound on Bayes error; and

optimize the objective function through gradient decent, wherein all dimensions of a matrix are optimized via optimizing the objective function;

wherein the optimizing is carried out ~~in an unconstrained manner~~ over all possible matrices; and

wherein the objective function is initialized using an LDA matrix;

upon convergence of the optimization, transform all the records into $y = \theta x$ to produce final features;

wherein the apparatus utilizes a processor to provide pattern recognition; and

wherein said pattern recognition is speech recognition.

8. (Cancelled)

9. (Cancelled)

10. (Original) The apparatus according to Claim 7, wherein said transformer is further adapted to query whether the optimized objective function converges.

11. (Original) The apparatus according to Claim 10, wherein said transformer is further adapted to repeat optimization of the objective function if the optimized objective function does not converge.

12. (Cancelled)

13. **(Currently Amended)** A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for providing pattern recognition, said method comprising the steps of:

inputting a speech pattern;

utilizing a processor to provide minimum Bayes error feature selection via
transforming the input pattern to provide a set of ~~at least one feature~~ features for a classifier which classifies into classes, wherein there is only one feature space transformation for all classes; and

providing final features to the classifier, wherein the classifier utilizes a processor to provide a final output classification result;

said transforming step comprising the step of directly minimizing the probability of subsequent misclassification in a projected space of the at least one feature ~~in the~~ classifier; said direct minimizing step comprising:

performing a full-covariance gaussian clustering of input records for every
class;

developing an objective function by way of means, covariances and priors,
wherein said objective function either:

maximizes an average pairwise divergence and relates it to the
Bayes error; or

directly minimizes an upper bound on Bayes error; and

optimizing the objective function through gradient decent, wherein all dimensions of a matrix are optimized via optimizing the objective function;

wherein the optimizing is carried out ~~in an unconstrained manner~~ over all possible matrices; and

wherein the objective function is initialized with an LDA matrix;

upon convergence of the optimization, transforming all the records into $y = \theta x$ to produce the at least one final feature;

wherein said pattern recognition is speech recognition.

14. **(Previously Presented)** The method according to claim 1, wherein said objective function is an average pairwise divergence related to the probability of misclassification of a projected space based on classes having uniform prior probabilities.

15. **(Cancelled)**

16. **(Cancelled)**

17. **(Previously Presented)** The method according to claim 1, wherein said objective function comprises means, covariances, and prior probabilities.

18. **(Previously Presented)** The method according to claim 1, wherein said objective

function is expressed by the following equation:

$$D_{\theta} = \frac{1}{C(C-1)} \text{trace} \left\{ \sum_{i=1}^C (\theta \Sigma_i \theta^T)^{-1} \theta S_i \theta^T \right\} - p$$

$$\text{where } S_i = \sum_{j \neq i} \Sigma_j + (\mu_i - \mu_j)(\mu_i - \mu_j)^T, i = 1, \dots, C.$$